# The Functional Uniform Prior
# An exploration towards its usefulness in non-linear (mixed) models

**Emmanuel Lesaffre**[1,2]
and
David Dejardin[3], Gwénaël Le Teuff[4], Mia Meyer[2], François Mercier[3], Paul Mostert[2], Shuting Tang[5]

[1]Interuniversity Institute for Biostatistics and statistical Bioinformatics, KU Leuven & University Hasselt, Belgium

[2]Department of Statistics and Actuarial Science, University of Stellenbosch, South Africa

[3]Product Development Data Sciences, F.Hoffmann-La Roche AG, Basel, Switzerland

[4]Oncostat U1018 Inserm, University Paris-Saclay, Villejuif, France

[5]Methodology & Statistics, Faculty of Psychology & Neuroscience, Maastricht University

**Bayes2025 meeting, Leiden — October 22, 2025**

# 1
# What are Functional Uniform Priors?

- Motivation to choose non-informative (read: vague) priors: import as little information into the analysis as possible

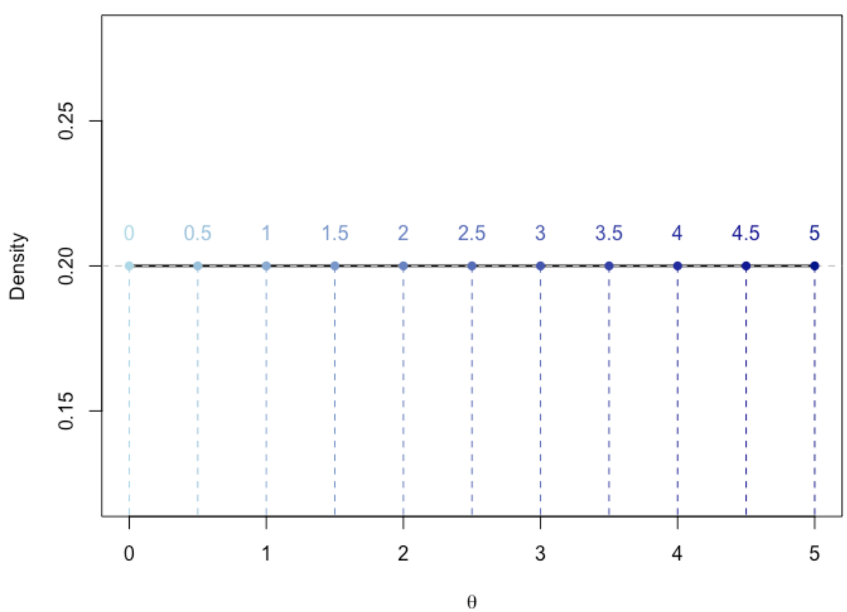
- Choice of non-informative priors:
  - Often (wide) uniform priors are taken on the (location) parameter space
  - Uniform priors may work well for linear and generalized linear (mixed) models
  - But, could be unintentionally informative for **non-linear (mixed) models**

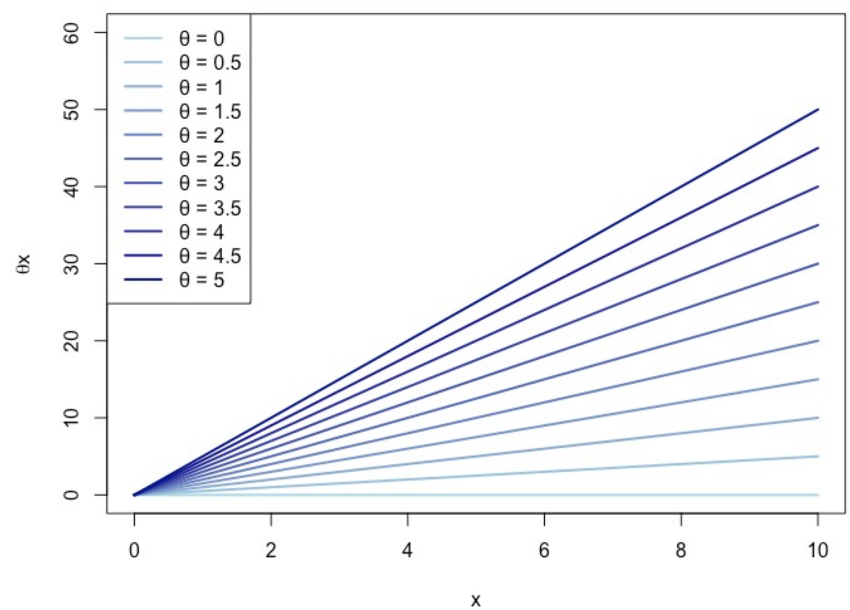
- Proposal of Bornkamp (2012, 2014):
  - Choose uniform prior on the (non-linear) functional space
  - He called such a prior: **Functional Uniform Prior**

# 2
# The problem

▷ Take a uniform prior for $\theta$ in [0, 5] for $\boldsymbol{y = \theta x}$

▷ Look at the corresponding linear functions

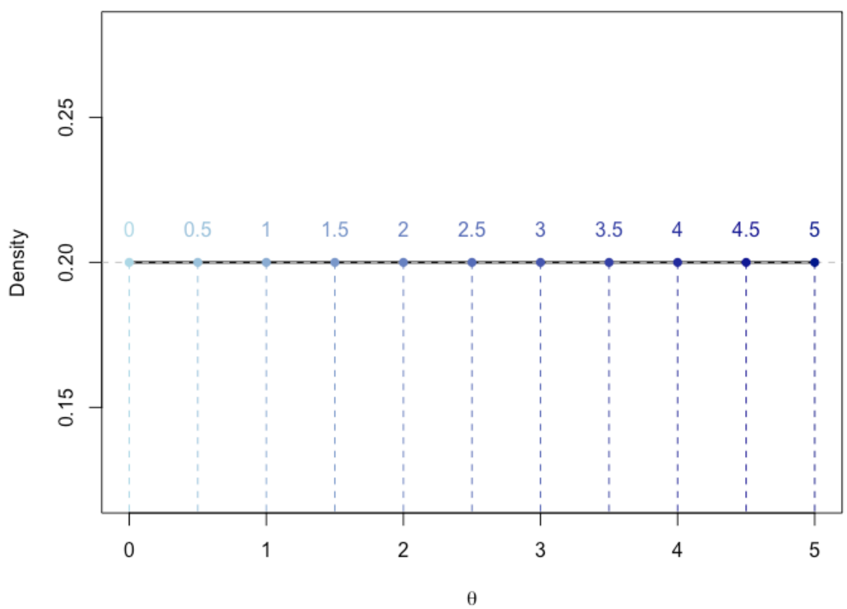


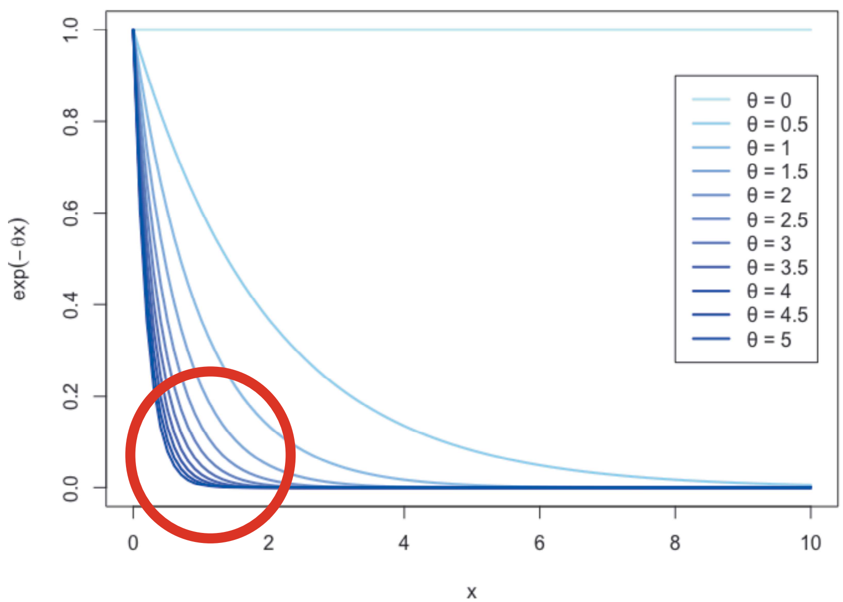




⇒ The straight lines are spread evenly in space

▷ Take again a uniform prior for $\theta$ in [0, 5] for $\boldsymbol{y = \exp(-\theta\, x)}$

▷ But now look at the corresponding nonlinear functions

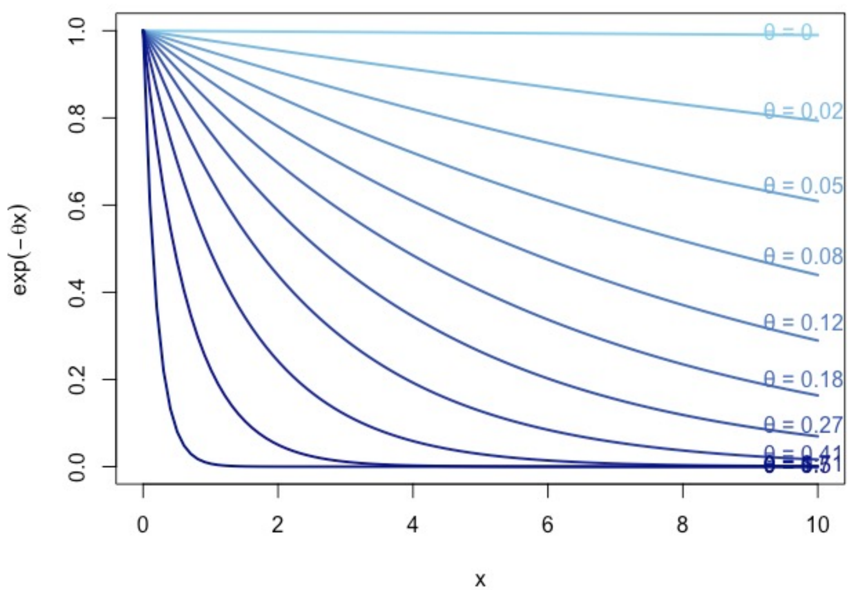


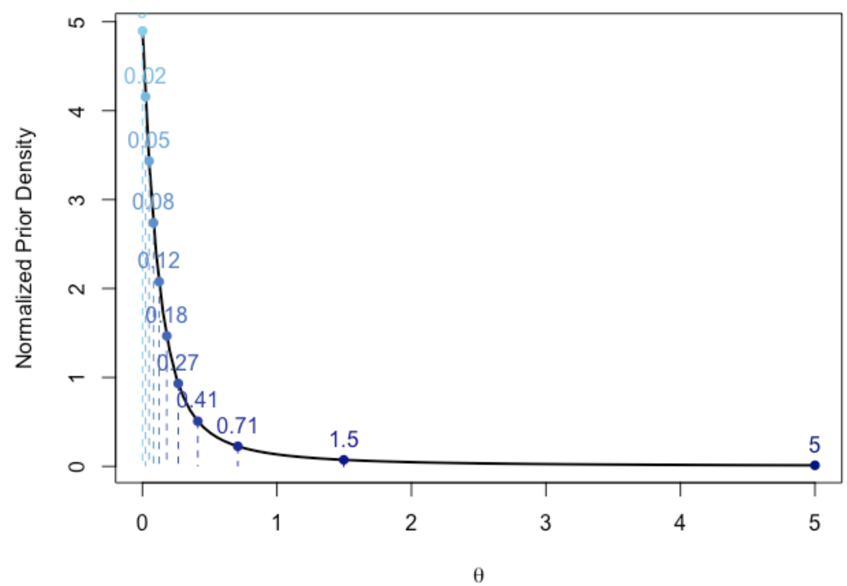




⇒ There is a high concentration of exponential functions near the axes

# 3
# Proposal of Bornkamp (2012)

- Prior should ensure that non-linear functions are **evenly spread in space according to some metric**
- Look at $y = \exp(-\theta\, x)$ for different values of $\theta$
- LHS: Evenly spread exponential functions, RHS: Functional Uniform Prior



# 4
# Construction of FUPs

- **Main idea**: prior ensures that regions in the parameter space corresponding to equally sized neighborhoods in the function space
- Suppose $\mu(x, \boldsymbol{\theta})$ represents a regression function:
  - $\mu(x, \boldsymbol{\theta})$ is nonlinear in the parameter vector $\boldsymbol{\theta}$
  - $x \in X \subset \mathbb{R}$ (range $X$ is typically range of interest for $x$)
  - $\boldsymbol{\theta} \in \boldsymbol{\Theta}$, a compact subspace of $\mathbb{R}^k$
- FUP:
  - Map the parameter $\boldsymbol{\theta}$ to the space of functions $\mu(x, \boldsymbol{\theta})$
  - Function space has **metric** $d(\mu(., \boldsymbol{\theta}), \mu(., \boldsymbol{\theta}'))$
  - Take a uniform distribution in this space
- Take e.g **Euclidean** $(L_2)$ **metric** $(v(dx)$ = measure on $X)$

$$d(\boldsymbol{\theta}, \boldsymbol{\theta}') = \sqrt{\int_X (\mu(x, \boldsymbol{\theta}) - \mu(x, \boldsymbol{\theta}'))^2 v(dx)}$$

▷ Approximate $\mu(x, \boldsymbol{\theta}) - \mu(x, \boldsymbol{\theta}')$ using a first-order Taylor expansion

▷ With $J_x(\boldsymbol{\theta}') = \frac{\partial \mu(x,\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}\Big|_{\boldsymbol{\theta}=\boldsymbol{\theta}'}$ and neglecting higher order term $O(\|\boldsymbol{\theta} - \boldsymbol{\theta}'\|^2)$:

$$d(\boldsymbol{\theta}, \boldsymbol{\theta}') \approx \sqrt{\int_X \left[J_x(\boldsymbol{\theta}')(\boldsymbol{\theta} - \boldsymbol{\theta}')\right]^2 v(dx)}.$$

▷ With $\mathbf{Z}^*(\boldsymbol{\theta}') = \int_X J_x(\boldsymbol{\theta}')^\top J_x(\boldsymbol{\theta}')\, v(dx)$ and $\mathbf{Z}^*(\boldsymbol{\theta}') \to \mathbf{Z}^*(\boldsymbol{\theta})$:

$$d(\boldsymbol{\theta}, \boldsymbol{\theta}') \approx \sqrt{(\boldsymbol{\theta} - \boldsymbol{\theta}')^\top \mathbf{Z}^*(\boldsymbol{\theta})(\boldsymbol{\theta} - \boldsymbol{\theta}')}.$$

▷ When $\mathbf{Z}^*(\boldsymbol{\theta})$ has finite positive eigenvalues, volume element in parameter space is $\sqrt{\det(\mathbf{Z}^*(\boldsymbol{\theta}))}\, d\boldsymbol{\theta}$. Assigning equal prior mass to equal metric volumes in the functional space yields the **functional uniform prior**:

$$p(\boldsymbol{\theta}) \propto \sqrt{\det(\mathbf{Z}^*(\boldsymbol{\theta}))}$$

▷ When $X$ is discrete, then integral is replaced by sum

# 5
# Comparison of FUP with Jeffreys' prior

- Both FUP and Jeffreys' prior enjoy **invariance properties**
- **Functional uniform prior**:

$$p(\boldsymbol{\theta}) \propto \sqrt{\det(\mathbf{Z}^*(\boldsymbol{\theta}))}$$

**Invariance property**: invariant under transformations that preserve distances between function shapes, and thus reparametrizations yield equivalent priors

- **Jeffreys' prior** (FUP for symmetric Kullback-Leibler distance):

$$p(\boldsymbol{\theta}) \propto \sqrt{\det(\mathbf{I}(\boldsymbol{\theta}))} \quad \text{with} \quad \mathrm{I}(\boldsymbol{\theta}) = \mathbb{E}\left[-\frac{\partial^2}{\partial\boldsymbol{\theta}^\top\partial\boldsymbol{\theta}}\log p(y \mid \boldsymbol{\theta})\right]$$

**Invariance property**: invariant under smooth reparameterizations in the parameter space

- Practical use:
  - **Jeffreys' prior**: depends on the chosen model, **and the design matrix** for regression models$\Rightarrow$ **may depend on the actually collected data**
  - **FUP**: depends on the chosen model **and the $X$-interval** $\Rightarrow$ **can be determined prior to data collection**

# 6
# Why is the FUP not more popular?

▷ The FUP

- elegant and nice concept that appears quite relevant for non-linear models
- BUT, only two publications have appeared: Bornkamp (2012, 2014)

⇒ **Why is the prior not popular?**

⇒ **Operating characteristics** of the FUP (frequentist evaluation)?

- Simulations & practical analysis

▷ We looked at:

- Various models: Emax, logistic, power & TGI
- Various priors: uniform, vague normal, vague exponential, Jeffreys & FUP
- Multiparameter case

▷ But we also extended the FUP prior to **non-linear mixed** models

## 7
# Replay of Bornkamp (2012)'s simulations for Emax model

▷ FUP showed for **Emax model** somewhat better OCs (Bornkamp, 2012)

$$\text{Emax model}: \; y = \theta_1 + \theta_2 \frac{x}{(x+\theta)}$$

▷ The FUP for $\theta \in [0.004, 6]$ and keeping $\theta_1$ and $\theta_2$ fixed:

$$p_{Emax}(\theta) \propto \sqrt{\frac{1}{(\theta+4)^3\theta}}$$

▷ Simulations Bornkamp (2012):

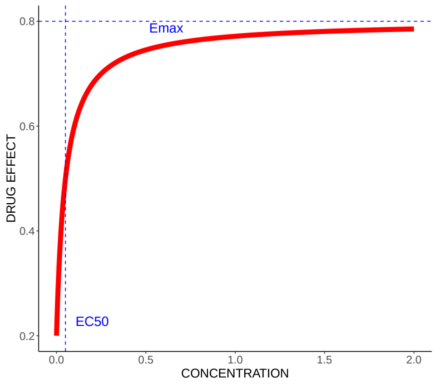
- giving $\theta$ a uniform, Jeffreys or FUP
- bias, average coverage prob and average CI length was somewhat better for FUP of $\theta$

▷ We repeated here the simulations (N=100) with

- $\theta_1 = 0.2$, $\theta_2 = 0.6$, $\theta = 0.05$
- n=15, 50, 100
- comparing operating characteristics for uniform, Jeffreys or FUP

▷ **Emax model**:

$$y = 0.2 + 0.6\,\frac{x}{(x+\theta)}, \quad \theta \in [0.004, 6]$$




▷ **Results**:

| Prior | Coverage 95% CI | Mean length 95% CI | MAE (mean) | MAE (median) |
|---|---|---|---|---|
| | **n = 15** | | | |
| Uniform | **0.971** | 0.5785 | 0.1678 | 0.1213 |
| Jeffreys | 0.987 | 0.4205 | 0.1094 | 0.0780 |
| FUP | 0.991 | **0.3763** | **0.0859** | **0.0644** |
| | **n = 50** | | | |
| Uniform | 0.972 | 0.1812 | 0.0423 | 0.0373 |
| Jeffreys | 0.975 | 0.1587 | 0.0329 | **0.0305** |
| FUP | **0.952** | **0.1527** | **0.0332** | 0.00329 |

▷ Overall FUP gives somewhat better results

# 8
# The Tumor Growth Inhibition (TGI) model

- **TGI model** is a **popular model in oncology**:
  - describes **Sum of Longest Tumor Diameters (SLD)** over time
  - captures both **tumor shrinkage** and **regrowth dynamics**
  - enables a more granular understanding of drug effects and disease progression
- **A typical TGI model** has two kinetic components:
  - **shrinkage rate**: quantifies drug-induced tumor reduction
  - **growth rate**: quantifies tumor progression in absence of treatment
- We considered:
  - Population TGI model
  - Mixed effects TGI model

▷ **Mathematical description of mixed effects TGI model**:

$$
\begin{aligned}
Y_{ij} &= f(t_{ij}, \boldsymbol{\psi}_i) + \epsilon_{ij} \\
&= y_{i0} \times \Big[\underbrace{\exp\left(-e^{\log k_s + b_{1i}} \cdot t_{ij}\right)}_{\textbf{shrinkage}} + \underbrace{\exp\left(e^{\log k_g + b_{2i}} \cdot t_{ij}\right)}_{\textbf{growth}} - 1\Big] + \epsilon_{ij}
\end{aligned}
$$

▷ where:

- $y_{i0}$: value of SLD at baseline
- $Y_{ij}$: SLD at time $t_{ij}$ for patient $i$
- $\log k_s$ and $\log k_g$ (in units of 1/year) represent population-level log (shrinkage & growth rate)
- $\epsilon_{ij} \sim \mathcal{N}(0, \sigma^2)$: additive residual error term
- random effects $b_{mi} \sim \mathcal{N}(0, \sigma^2_{bm})$ $(m = 1, 2)$ account for individual deviations

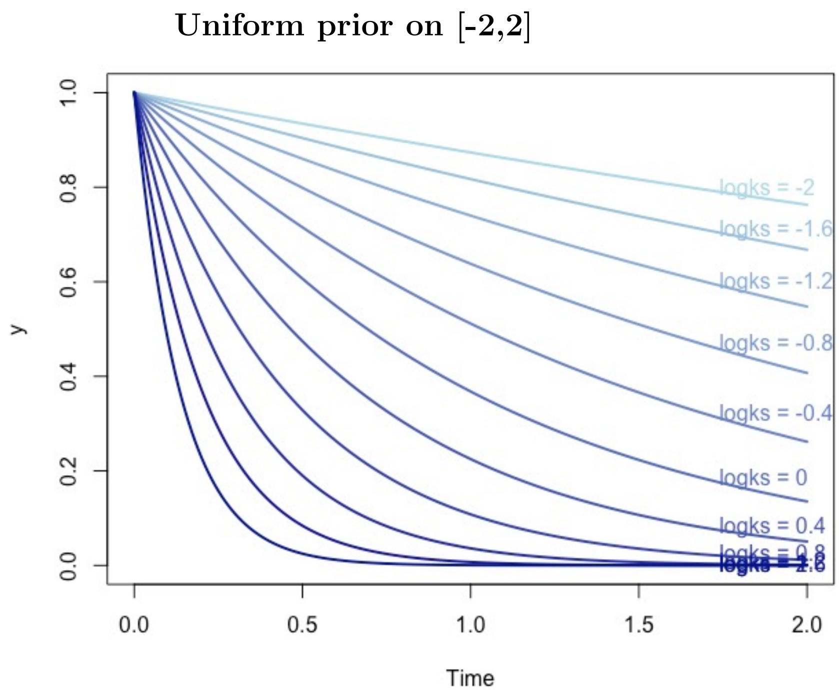
▷ The structure allows the model to accommodate both population-level trends and individual-level heterogeneity in tumor dynamics

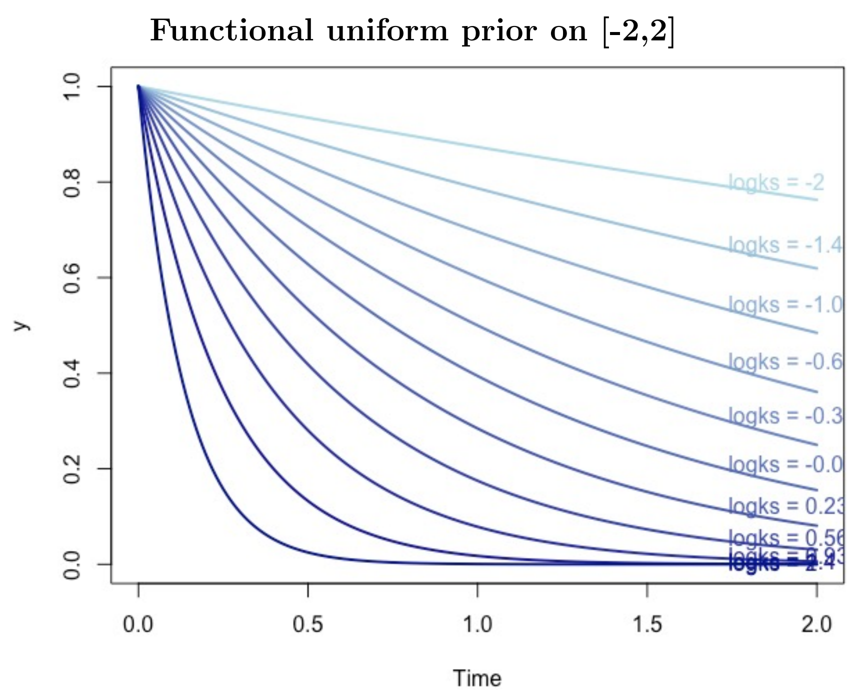
▷ **We applied the FUP construction to the fixed effects part!**

## **FUP for shrinkage rate component**:

- ▷ Shrinkage rate component: $\mu(t, \log k_s) = \exp(-e^{\log k_s} \cdot t), \quad t \in T = (0, 2)$
- ▷ Limits of $k_s \in [0.14, 7.4] \Rightarrow \log k_s \in [-2, 2]$ (based on literature search)

$$p(\log k_s) \propto \Big(\frac{1}{4}\exp(-\log k_s) - 2\exp(\log k_s - 4\exp(\log k_s)) - \exp(-4\exp(\log k_s)) \\ -\frac{1}{4}\exp(-\log k_s - 4\exp(\log k_s))\Big)^{1/2}$$

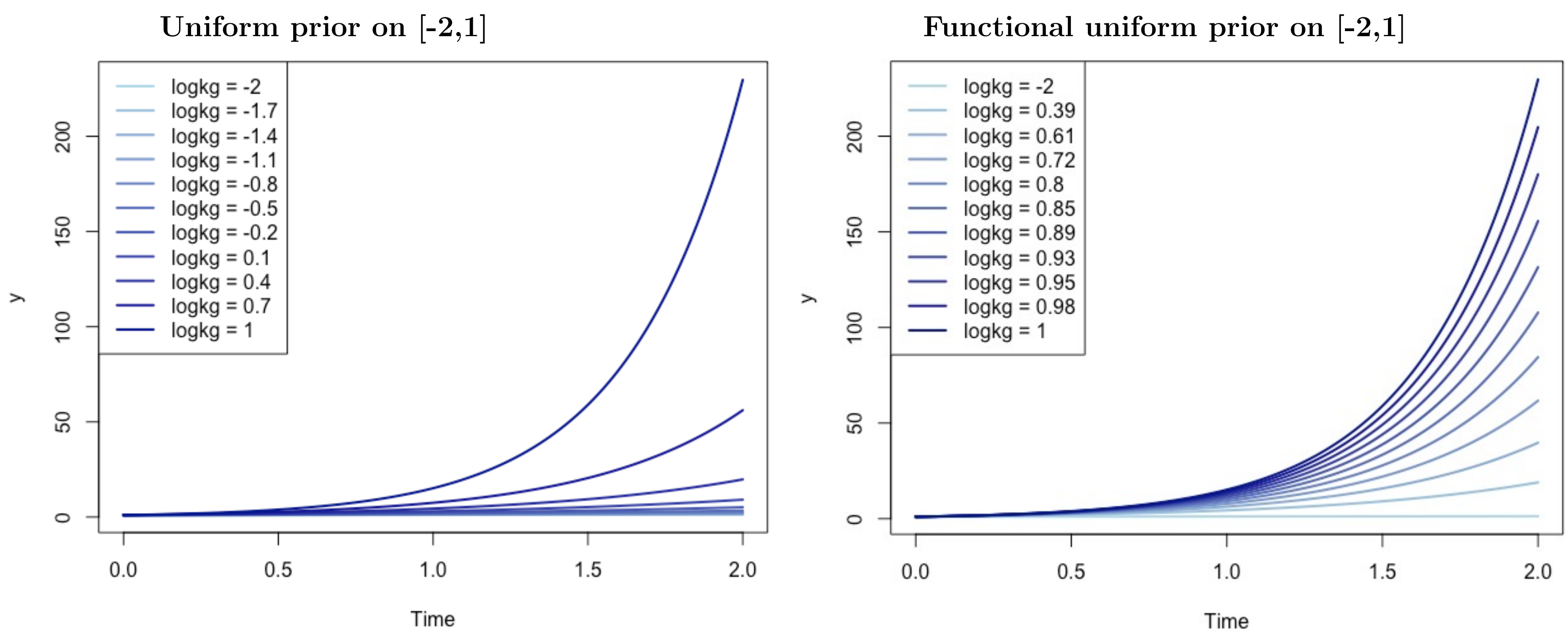


## **FUP for growth rate component**:

- ▷ Growth rate component: $\mu(t, \log k_g) = \exp(e^{\log k_g} \cdot t), \quad t \in T = (0, 2)$
- ▷ Limits of $k_g \in [0.14, 2.72] \Rightarrow \log k_g \in [-2, 1]$ (based on literature search)

$$p(\log k_g) \propto \Big(-\frac{1}{4}\exp(-\log k_g) + 2\exp(\log k_g + 4\exp(\log k_g)) - \exp(4\exp(\log k_g)) + \frac{1}{4}\exp(-\log k_g + 4\exp(\log k_g))\Big)^{1/2}$$



## **FUP for the two components jointly**:

- Bivariate function:

$$\mu(t; \log k_s, \log k_g) = \exp(-e^{\log k_s} \cdot t) + \exp(e^{\log k_g} \cdot t) - 1, \quad t \in T = (0, 2)$$

$$p(\boldsymbol{\theta}) \propto \sqrt{\det\left(\int_T J_t(\boldsymbol{\theta})^\top J_t(\boldsymbol{\theta})\, dt\right)}$$

where $J_t(\boldsymbol{\theta})$ is the Jacobian of $\mu(t; \boldsymbol{\theta})$ with respect to $\boldsymbol{\theta}$, evaluated at time $t$.

- The joint FUP has similar OCs as the combination of separate FUPs

- Note:

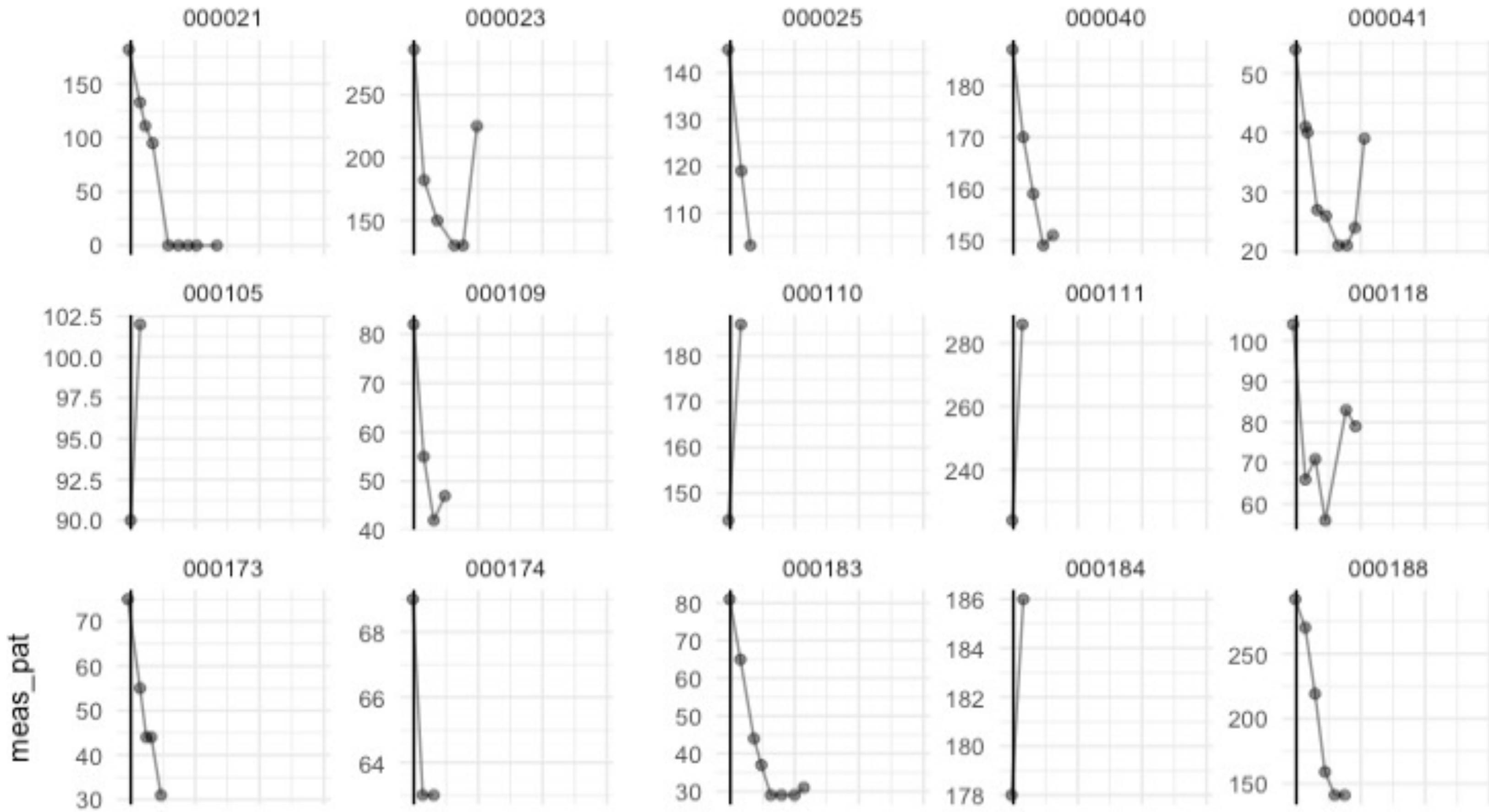
$\triangleright$ $\exp(e^{\log k_g} \cdot t) = \exp(-e^{\log k_s} \cdot -t)$

$\Rightarrow$ Shrinkage and growth part are each other's mirror with opposite X-interval

# 9
# Application of TGI model to PRIME data

- **PRIME data set**:
  - ▷ Douillard et al. (2010) Randomized, Phase III Trial of Panitumumab With Infusional Fluorouracil, Leucovorin, and Oxaliplatin (FOLFOX4) Versus FOLFOX4 Alone As First-Line Treatment in Patients With Previously Untreated Metastatic Colorectal Cancer: The PRIME Study. J of Clin Oncology
  - ▷ Open-label, multicenter, phase III trial that compared the efficacy of panitumumab-FOLFOX4 with FOLFOX4 alone in 1,183 patients with previously untreated mCRC according to tumor KRAS status
  - ▷ Patients were randomly assigned 1:1 to receive either panitumumab-FOLFOX4 or FOLFOX4.
  - ▷ An initiative made certain available for use to the public, see `https://data.projectdatasphere.org/projectdatasphere/html/access`
  - ▷ Analysis is done on **50 patients from the control group**
- Software: NIMBLE 1.3 and Stan 2.37 (two Master students worked on it)

## **PRIME data set**: Sum of Longest Tumor Diameters of 15 randomly chosen patients from control group



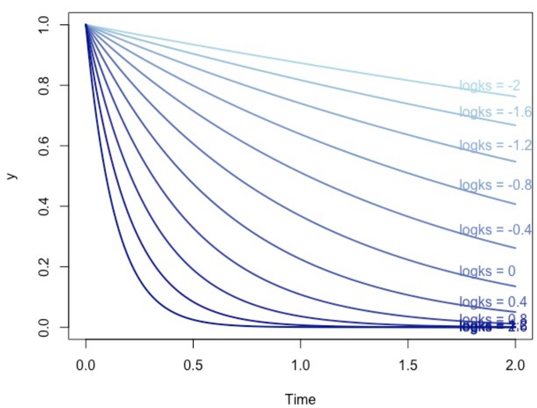
# Result based on different priors

| Settings | Mean | Median | Var | 95% | Width | IMSE |
|---|---|---|---|---|---|---|
| $\log k_s$ | | | | | | |
| 2: Vague uniform | 0.43 | 0.43 | 0.03 | [0.05, 0.77] | 0.72 | 109.12 |
| 4: Separate FUP | 0.42 | 0.43 | 0.03 | [0.05, 0.74] | 0.69 | 109.59 |
| $\log k_g$ | | | | | | |
| 2: Vague uniform | -1.37 | -1.37 | 0.04 | [-1.80, -0.98] | 0.82 | 109.12 |
| 4: Separate FUP | -1.32 | -1.31 | 0.04 | [-1.73, -0.96] | 0.77 | 109.59 |

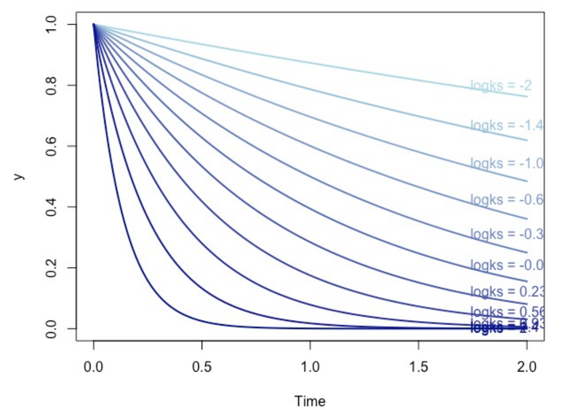
- ▷ **Observation: FUP gives quite similar results as uniform prior** How come?
- ▷ Some (possible) explanations:
  - FUP is similar to uniform on shrinkage part ⇒ not much different effect expected
  - FUP is quite different on growth part, but **PRIME data on growth is sparse** ⇒ not much different effect expected
  - Uniform prior was taken on restrictive range, inspired by literature results ⇒ **informative?**
  - Are oncology statisticians too informative or too smart?

# Shrinkage part

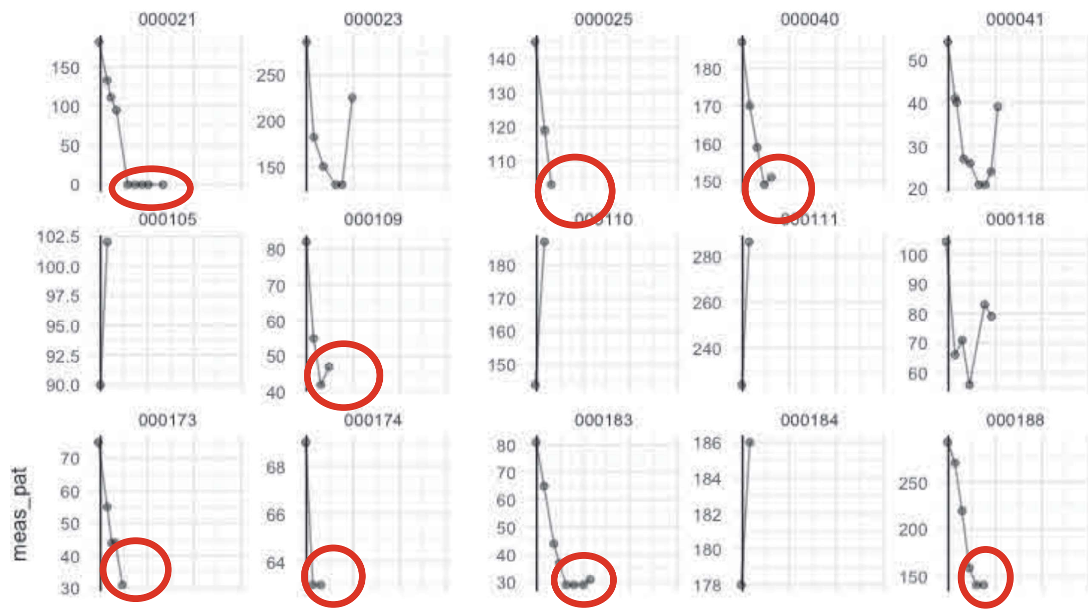
**Uniform prior on [-2,2]**




**Functional uniform prior on [-2,2]**

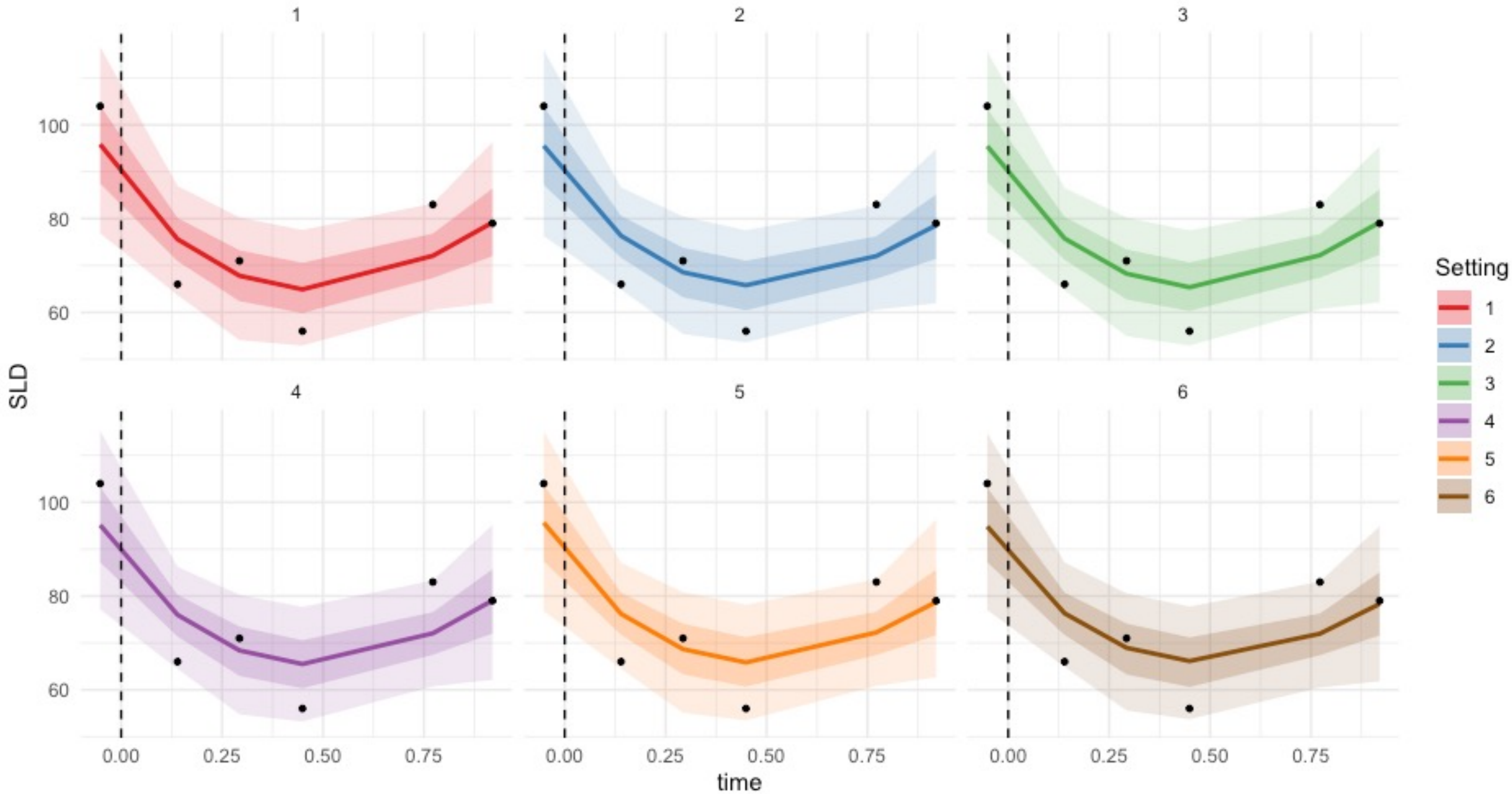



# Absence of growth data



# Comparison of predicted Sum of Longest Tumor Diameters

For a randomly selected patient




1=vague normal, 2=vague uniform, 3=vague exponential
4=separate FUP (ana), 5=separate FUP (num), 6=joint FUP (num)

## 10
## Simulation studies

▷ With simulations **one can perhaps chose better scenarios**

▷ **Model**:

$$\begin{aligned} Y_{ij} &= f(t_{ij}, \boldsymbol{\psi}_i) + \epsilon_{ij} \\ &= y_{i0} \times [\underbrace{\exp\left(-e^{\log k_s + b_{1i}} \cdot t_{ij}\right)}_{\textbf{shrinkage}} + \underbrace{\exp\left(e^{\log k_g + b_{2i}} \cdot t_{ij}\right)}_{\textbf{growth}} - 1] + \epsilon_{ij} \end{aligned}$$

**Several TGI scenarios**:

- Fixed effects and mixed effects TGI model
- Priors: vague normal, uniform, exponential, separate and joint FUP
- Sample sizes: 15, 50, 100
- 5 fixed time points
- Number of simulations: 100

+ Focus on growth part of the model (largest difference FUP with uniform)

## Results simulation study

- ▷ FUP does not give consistently better results compared to other vague priors
- ▷ Joint FUP gives similar results as separate FUPs
- ▷ Extending the interval of the parameters (more clustering towards X-axis) has no effect
- ▷ Extending the length of the growth part does not have an effect

# 11
# (Tentative) Conclusions

- ▷ FUP is based on an elegant mathematical/statistical concept
- ▷ FUP has the practical advantage that it can be computed prior to having collected the data, in contrast to Jeffreys' prior
- ▷ FUP can be seen as a generalization of Jeffreys' prior
- ▷ But:
  - ◦ Practical advantage of a FUP seems to depend on the chosen model and the availability of data
- ▷ Probably further research is needed ...

**ANY IDEAS FROM THE FLOOR?**

# Back-up slides

# 12
# FUP for Logistic Growth Model

- Logistic Growth Model:
  - ▷ describes evolution of quantity over time, e.g. tumor volume, bacterial population, ...
  - ▷ assumes growth initially accelerates, then decelerates to stabilize at maximum level

$$f(t;\boldsymbol{\theta}) = \frac{K}{1+\exp[-r\,(t-t_0)]} \qquad \boldsymbol{\theta} = (K, r, t_0)$$

  - ∘ $K > 0$: *carrying capacity*, i.e., the maximal level the curve approaches as $t \to \infty$
  - ∘ $r > 0$: *growth rate*, controlling how steeply the curve increases
  - ∘ $t_0$: *inflection time*, time at which growth switches from accelerating to decelerating and $f(t = t_0) = K/2$

- Note: $r$ and $t_0$ may be highly correlated: different combinations of $(r, t_0)$ may produce similar curve shapes around the inflection region

## Joint Functional Uniform Prior for $\boldsymbol{\theta} = (r, t_0)$

▷ $K$ enters the mean function linearly

⇒ FUP only for $\boldsymbol{\theta} = (r, t_0)$

▷ Logistic growth mean function

$$\mu(t; \boldsymbol{\theta}) = \frac{1}{1 + \exp[-r(t - t_0)]} \qquad t \in T = \{0, 1, 2, 3, 5, 7, 10, 14, 21, 28\}$$

▷ Using $L^2$ metric on functions, at each $t \in T$

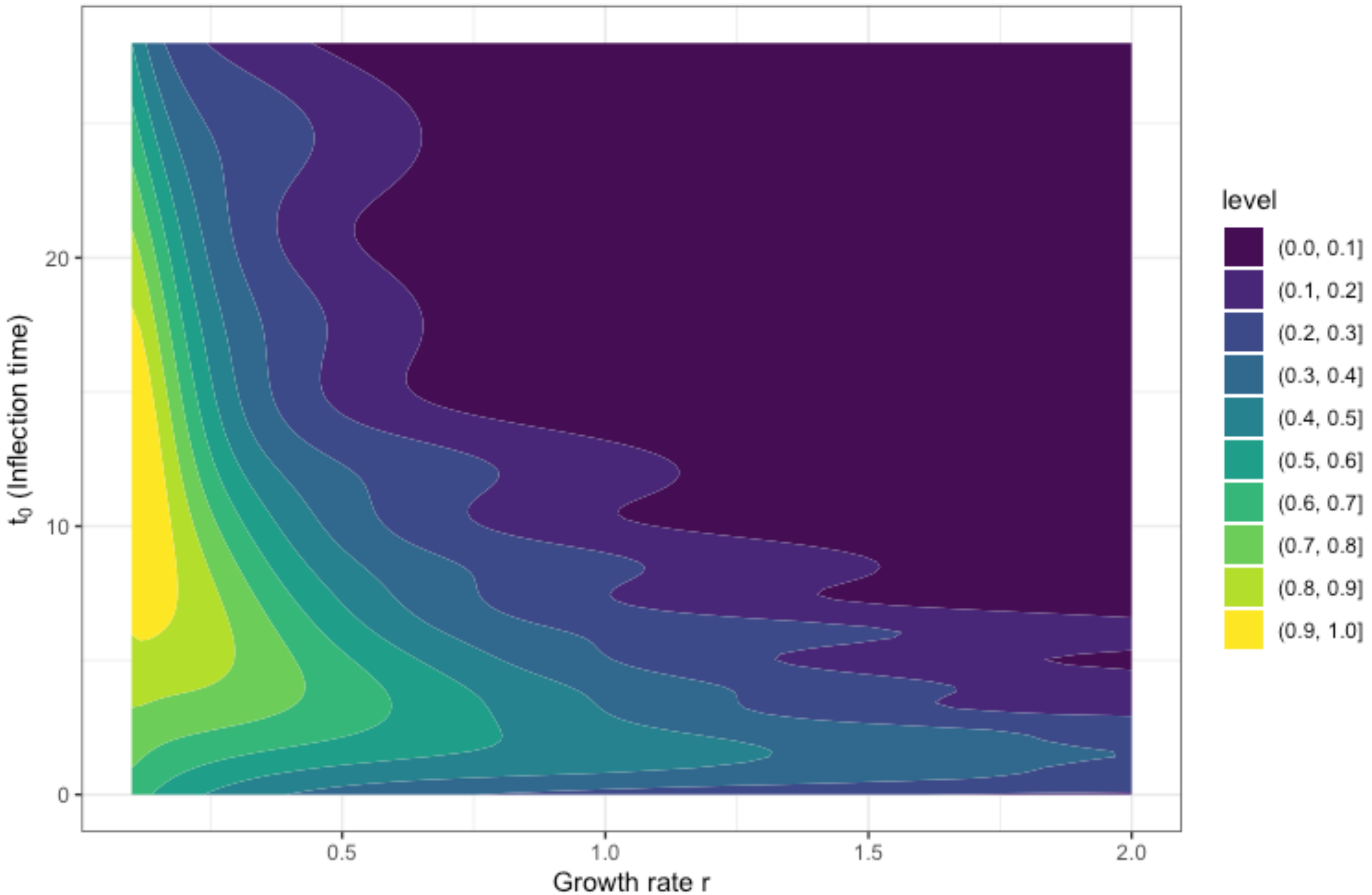
$$J_t(\boldsymbol{\theta})^\top J_t(\boldsymbol{\theta}) = \left( \frac{\exp[-r(t - t_0)]}{\left(1 + \exp[-r(t - t_0)]\right)^2} \right)^2 \begin{bmatrix} (t - t_0)^2 & -r(t - t_0) \\ -r(t - t_0) & r^2 \end{bmatrix}$$

▷ Summing for discrete $t$ over $t \in T$ yields the $2 \times 2$ matrix

$$\mathbf{Z}(\boldsymbol{\theta}) = \sum_{t \in T} J_t(\boldsymbol{\theta})^\top J_t(\boldsymbol{\theta})$$

⇒ Joint FUP: $p(\boldsymbol{\theta}) \propto \sqrt{\det\left(\mathbf{Z}(\boldsymbol{\theta})\right)}$

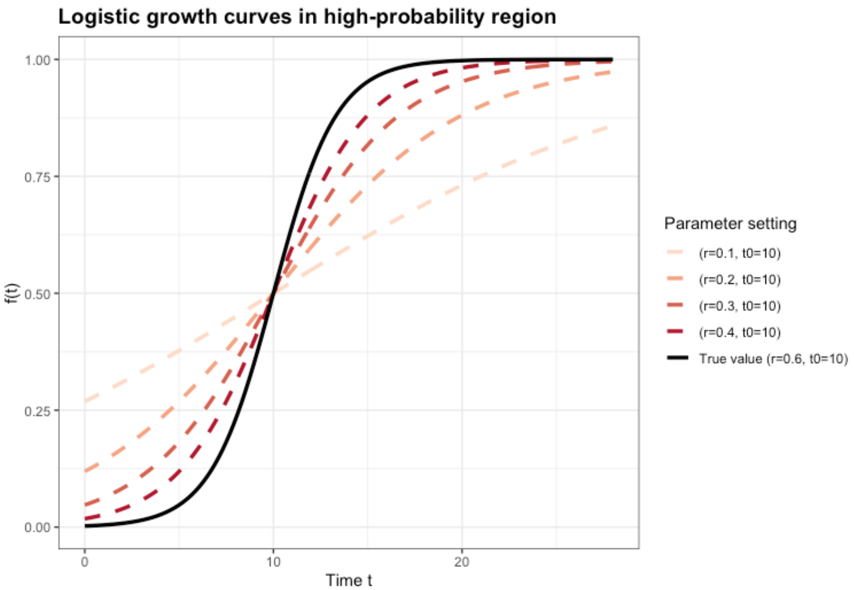
# Joint Functional Uniform Prior for $\boldsymbol{\theta} = (r, t_0)$

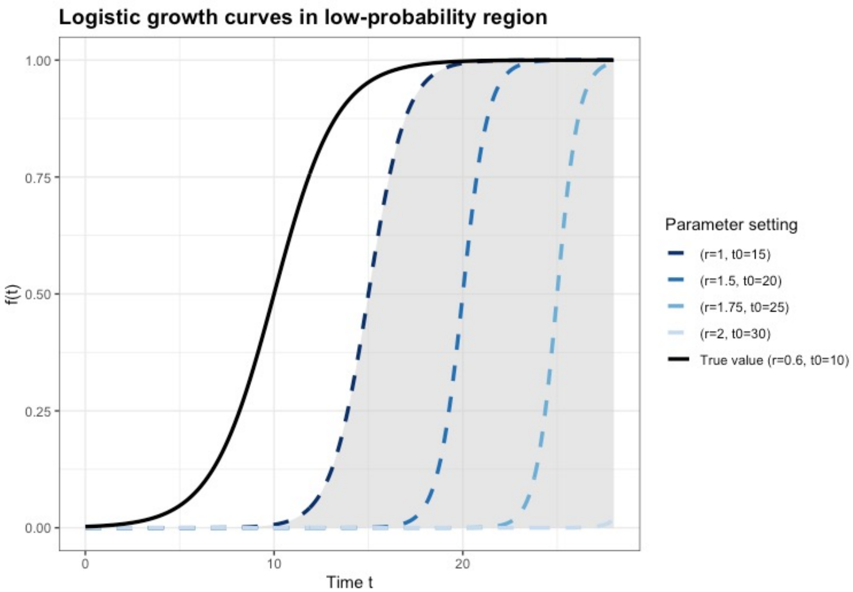



High-density regions (yellow): distinct functional shapes

Low-density regions (dark blue): nearly indistinguishable curves

# High- and low probability regions



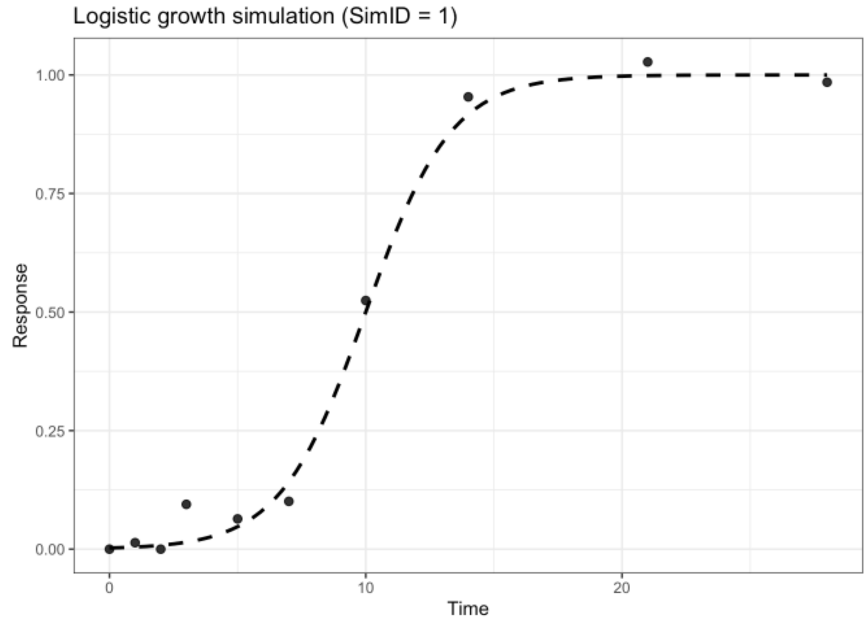



## Simulation set up

▷ Simulation experiment:

- $\boldsymbol{\theta}_{\text{true}} = (K, r, t_0) = (1.0,\ 0.6,\ 10), \quad \sigma_{\text{true}} = 0.05$
- $t \in T = \{0, 1, 2, 3, 5, 7, 10, 14, 21, 28\}$
- $Y_i \sim \mathcal{N}(\mu(t_i; \theta_{\text{true}}),\ \sigma^2_{\text{true}})|^+, \quad i = 1, \ldots, N$

▷ Example simulated data:



▷ Two prior settings:

- ▷ **Uniform priors (U):**
  $K \sim \mathsf{U}(0.5, 2.0), \quad r \sim \mathsf{U}(0.05, 2.0), \quad t_0 \sim \mathsf{U}(-2, 30), \quad \sigma \sim \mathsf{U}(0, 0.5)$
- ▷ **Functional uniform prior (FUP):**
  Joint FUP on $(r, t_0)$: $\pi_{\mathsf{FUP}}(r, t_0) \propto \sqrt{\det(\sum_{t \in T} J_t(\boldsymbol{\theta})^\top J_t(\boldsymbol{\theta}))}$
  and
  $K \sim \mathsf{U}(0.5, 2.0), \quad \sigma \sim \mathsf{U}(0, 0.5)$

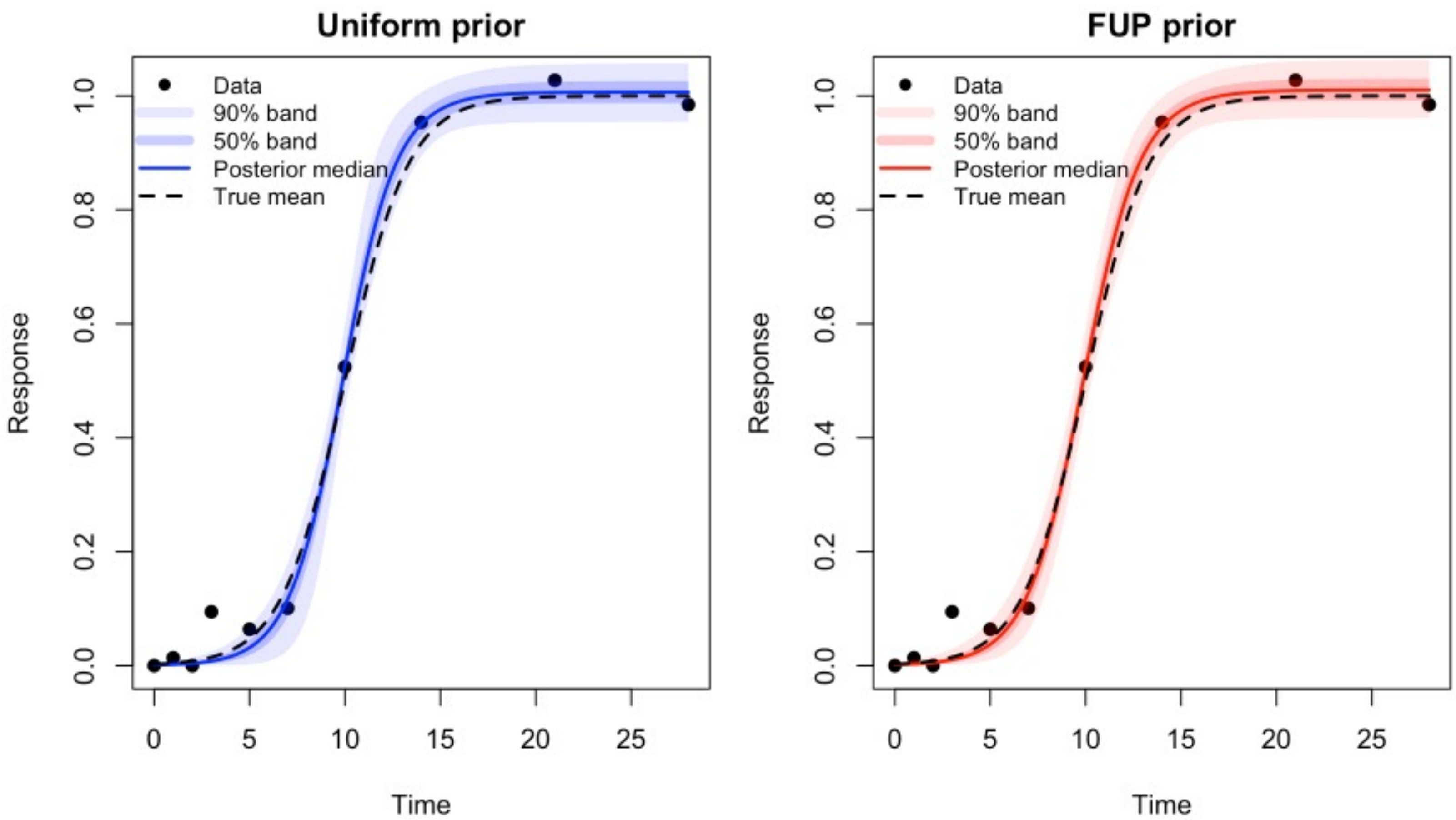
- MCMC in `R` using **Nimble** package & MH sampler
- 3 independent chains in parallel, each with 50,000 iterations
- Burn-in of 5,000 iterations, and thinning of 10
- FUP setting: $\log \pi_{\mathsf{FUP}}(r, t_0)$ precomputed on a grid ($r \in [0.05, 2.0]$, $t_0 \in [-2, 30]$), interpolated bilinearly during MCMC using a Poisson zero-trick formulation

# Simulation results

| Setting | Mean | Bias | R.Bias | CP95 | MIMSE |
|---|---|---|---|---|---|
| **K** | | | | | |
| 1 (Uniform) | 1.0010 | 0.0010 | 0.0010 | 0.88 | 0.0756 |
| 2 (FUP) | 1.0056 | 0.0056 | 0.0056 | 0.89 | 0.0738 |
| **r** | | | | | |
| 1 (Uniform) | 0.6388 | 0.0388 | 0.0647 | 0.86 | 0.0756 |
| 2 (FUP) | 0.6019 | 0.0019 | 0.0032 | 0.86 | 0.0738 |
| $t_0$ | | | | | |
| 1 (Uniform) | 9.996 | -0.0044 | -0.00044 | 0.92 | 0.0756 |
| 2 (FUP) | 10.018 | 0.0178 | 0.00178 | 0.92 | 0.0738 |

R.Bias: relative bias

MIMSE: Mean Integrated Mean Squared Error.


Uniform prior
FUP prior
Data
90% band
50% band
Posterior median
True mean
Response
Time
0.0
0.2
0.4
0.6
0.8
1.0
0
5
10
15
20
25

## Tentative conclusions (again)

- ▷ FUP is based on an elegant mathematical/statistical concept
- ▷ FUP puts little prior on areas of $(r, t_0)$ which indistinguishable curves
- ▷ But again: practical advantage is little